25th ANNUAL NATIONAL CONFERENCE ON MANAGING ENVIRONMENTAL QUALITY SYSTEMS

APRIL 24-27, 2006

Marriott Renaissance, Austin, Texas

Technical Papers

Detection and Reporting Limits

- S.Wasson, Method Detection Limits and Reporting Non-Detects in Microbiology - 1:00 PM
- C.Davis, When Laboratorires Should Not Censor Data - 1:30 PM
- J. Martin, New Statistical Procedure for Determining MRL Under the UCMR2 - 2:00 PM

Method Detection Limits and Non-Detects in the World of Microbiology

Shirley J. Wasson
Air Pollution Prevention and Control Division
National Risk Management Research Laboratory
United States Environmental Protection Agency
Research Triangle Park, NC 27711

Abstract
Examining indoor air for microorganisms is generally performed by sampling for viable microbes, growing them on sterile media under ideal conditions, and counting the colony forming units (CFUs). A negative result does not indicate that the source of the sample was free of fungi or bacteria, however, only that if present, the number of viable fungi or bacteria was below the limits of detection of the test.

This situation is problematic where government officials declare a building inhabitable on the basis of "no growth in any environmental sample". The Government Accountability Office (GAO) and the public want the answer to a more pragmatic question: "Is this building contaminated?" Likewise lack of confidence in negative results is at issue where professional remediators of water damaged buildings are trying to gauge efficacy of their efforts to eradicate fungi and their spores, or where in-duct ultraviolet light technologies are being tested for kill of vegetative bacteria and the manufacturers wish to publish efficiency ratings for the devices.

The dichotomy is that decision makers and building occupiers want to know with certainty whether dangerous or potentially dangerous microbiological organisms still exist in the structure while even the most sophisticated sampling and analysis methods available cannot provide conclusions with 100% certainty. Enter here government risk assessors who must grapple with the problem using the same data.

This paper explores sampling indoor air and surfaces for microorganisms, their analysis by conventional and state of the art methods, the interpretation of the results, and the state of governmental regulation of acceptable levels of such organisms and their effluents.

Introduction

The U.S. Environmental Protection Agency (EPA) defines the method detection limit (MDL) as the lowest amount that differentiates a sample that contains the substance from one that does not, and the quantification limit as the lowest amount of a substance that can be measured with a stated level of confidence (40 CFR 1984). In chemistry, it is understood that non-detects may contain some of the analyzed substance but in a small enough quantity that it can be considered zero and of little concern. Government risk

assessors evaluating risk of highly toxic substances may require the development of more sensitive sampling and analytical methods.

The traditional thought processes regarding MDLs are strained, however, when measuring microscopic organisms such as vegetative bacteria, fungi, spores, and viruses. Traditional methods for analyzing microbes are so sensitive that they can detect even one viable entity, that is, one capable of reproducing under ideal growth conditions for the organism, but several issues remain. Air samplers, for instance, have to be sited correctly to capture any viable organisms. Sampling methods can kill organisms via desiccation or otherwise fragment them and render them incapable of growth and therefore detection by viability methods, yet still exist in the environment as dangerous allergens. Analysis methods are based on morphological characteristics and can be inaccurate. Fragments and products of microorganisms, such as mycotoxins, must be sampled and identified another way. Any sampling and analysis plan must provide for both viable and non-viable microbes to be representative of what is really there.

Three Recent Scenarios That Illustrate the Problem

Determining residual biological threat agent following cleanup from a terrorist attack
In September and October 2001, letters containing the spores of *Bacillus anthracis* (anthrax) sent through the U.S. mail were opened by the staff in the offices of two U.S. senators and members of the media. A total of 22 persons contracted anthrax disease and five died. The U.S. Postal Service (USPS), the Centers for Disease Control and Prevention (CDC), and the EPA performed several sample collection and analytical activities in postal facilities in 2001 for the purpose of detecting anthrax. The sampling strategy was to target the most likely areas where anthrax might be found. On April 5, 2005, 3 ½ years later, in its testimony before the House of Representatives Subcommittee on National Security, Emerging Threats, and International Relations, the U.S. Government Accountability Office (GAO 2005) stated that probability sampling would have been a better choice. The GAO believes that probability sampling would have allowed decision makers to determine whether a building is contaminated with some defined level of confidence *even when all results are negative*. GAO wants validated sampling. Even now, 4 ½ years after the bioterrorist incident, sampling and analysis strategies are still at issue.

Determining residual fungi from cleanup after a natural disaster
On August 29, 2005, hurricane Katrina brought category 4 force winds and flood damage to several hundred square miles of the Louisiana and Mississippi Gulf Coast. Several levees were breached flooding up to 80% of the city of New Orleans and large areas of surrounding parishes. The U.S. Army Corp of Engineers declared the flooded areas "unwatered" on October 11, 2005, but the aftermath of soaked buildings and furnishings has brought a legacy of mold proliferation on a scale rarely seen in U.S. history. For those homes that are deemed remediable, a huge cleanup effort is being undertaken. Little of this work can be performed by the residential owners themselves because the microbial contamination is so extensive. For surface areas of 10 square feet or more affected by mold, EPA recommends professional cleanup. No sampling is recommended

for areas with visual mold growth, but during the cleanup professional remediators must perform to standards that bring the property to a condition at least as good as existed before the flood. Any such assurance will involve post-remedial sampling for residuals from fungi. What sampling and analysis plan will provide such confidence?

Determining residual bioaerosols after treatment with an efficient technology
Manufacturers are interested in verifying the airborne inactivation efficiency of their in-duct ultraviolet (UV) light air cleaning systems on the culturable challenge bioaerosol in the air circulating in heating, ventilation and air conditioning (HVAC) systems. In a test based on the inert particulate testing method for filters that is the basis for the American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE 1999) Standard 52.2, the testing organization tests a full-scale module of each UV device by installing it in a standard full-scale test duct that meets the ASHRAE standard. The selected challenge aerosol is injected into the inlet air stream upstream of a mixing baffle. Bioaerosol concentration is measured both downstream and upstream of the device to obtain the ratio of the surviving concentration to the challenge concentration of viable test organisms. Challenges with spore-forming bacteria such as *Bacillus atrophaeus* or a virus such as MS2 provide calculable efficiencies since they have measurable downstream survivors. Vegetative bacteria such as *Serratia marcescens*, however, are easy to kill with UV light and therefore often have *no measurable survivors downstream of the device*. How should efficiency for such bioaerosols be reported?

Solutions: State of the Science of Decontamination after a Terrorist Attack

At an EPA National Homeland Security Research Center Workshop (Dun et al 2005), Kenneth Martinez of the Centers for Disease Control (CDC) stated that the purposes for environmental sampling after a terrorist attack include assessing the nature and extent of contamination, identifying the sources, supporting risk assessment and public health decisions, and guiding re-occupancy decisions. The three sampling stages of a response are screening, characterization, and restoration. Workshop participants stated that once a deadly agent like *B. anthracis* has been identified, no further characterization is necessary until after building fumigation. During a decontamination event, a disinfectant must reach a specific dose (ppm-hours) to ensure efficacy. Post-fumigation sampling is performed to determine whether the dose has been delivered and the contaminating agent has presumably been eradicated. Some fumigators use biological indicators such as spores on steel coupons or paper strips, but the question remains how indicative are biological indicators of the condition of real world materials such as carpet, ceiling tiles, wood, painted walls, or fabric.

Standards for sampling decontaminated facilities are currently not available. CDC issued guidance (CDC 2002) on collecting environmental samples for culturing *B. anthracis* and established the Laboratory Response Network (LRN) to investigate and validate sampling and analytical methods for biological contaminants focusing on efficiency of surface sampling, air sampling, methods comparison, and variability. Research at the Edgewood Chemical Biological Center (ECBC) has been able to relate differences in kill efficacy to differences in surfaces, accounting for surface variability in real-world situations. The

Department of Defense (DOD) is establishing an environmental LRN similar to that of the CDC to harmonize sampling.

Paula Krauter of Lawrence Livermore National Laboratory (LLNL) presented research on developing a 15-hour method for processing biological indicator strips using a real-time polymerase chain reaction (PCR) technique called a rapid viability test protocol (RVTP) and compared it to a standard culture method which requires 7 days for results. The results were comparable except that the standard culture method reported a 1.5 % false positive rate while *the RVTP reported no false positives or negatives*.

Correct statistical design considers risk. The National Academy of Sciences (NAS 2005) addressed the question "How Clean is Safe?" in its publication "Reopening Public Facilities after a Biological Attack: A Decision–Making Framework." The NAS states that risk analysis informs interested parties of the probability of having any residual organisms in the building and of those residual organisms causing an infection in a human occupant, based on the detection limit, sampling efficiency, and dose-response data. They recommend convening an expert group and an Operations Working Group (OWG) composed of stakeholders, building managers, and decision makers to determine acceptable risk and whether a building can be declared safe for occupancy. If the risks cannot be determined with confidence because of high uncertainties associated with sampling or decontamination methods, the acceptable choice is to further decontaminate to increase the probability that the building is safe.

Decontamination of Biologically Contaminated Sites after a Flood Event

The issues of sampling and analysis of fungal microorganisms after a flood event are different from those after a terrorist attack. Generally a visual inspection leaves no doubt of the presence of mold, thus usually no sampling is indicated until after remediation work is completed. At that time the question then becomes what sampling and analysis plan will assure that the site is sufficiently remediated to occupy. Federal guidance provided by EPA's "Mold Remediation in Schools and Commercial Buildings" (USEPA 2001) and CDC's "Mold Prevention Strategies and Possible Health Effects in the Aftermath of Hurricanes Katrina and Rita" (CDC 2005) mainly defaults to industry experts.

The Institute of Inspection, Cleaning & Restoration Certification guide (IICRC 2003) defines indoor environmental conditions. A Condition 1 (normal fungal ecology) site may have settled spores, fungal fragments, or traces of actual growth whose identities, locations, and quantities are reflective of a normal fungal ecology for a similar indoor environment. In Condition 2 (settled spores) the site is primarily contaminated with settled spores that were dispersed directly or indirectly from a Condition 3 area, and may have traces of actual growth. A Condition 3 (actual growth) site is contaminated with the presence of actual mold growth and associated spores that may be active or dormant, visible or hidden. The IICRC recommends physical removal as the primary means of remediation to return the indoor environment to Condition 1 status and maintains that attempts to simply kill or encapsulate mold are not generally adequate. Physical removal

also has the advantage that it mitigates fungal toxins (mycotoxins) and/or spore wall components (glucans) which are the most likely etiology of building related allergy/immunological complaints.

The IICRC guidance relies on Indoor Environmental Professionals (IEPs) as third party inspectors to assess and declare a building returned to Condition 1. The IEPs may consult the American Conference of Governmental Industrial Hygienists guide (Macher 2005) for such assessment. Testing after mold remediation focuses on checking for removal of the water source and preparing an adequate sampling and analysis plan to capture the fungal ecology of the remediated space and comparing it to the "normal" fungal ecology for a similar indoor environment or to that immediately outdoors. If the remediated levels are higher or if the mold concentrations and species are significantly different, further remediation is required. In the absence of federal standards or recommendations for acceptable indoor airborne levels of viable mold, Baxter et al 2005 have attempted to devise criteria for differentiating "clean" from "moldy." Less than 1200 spores/m^3, <750 *Aspergillus/Penicillium* /m^3, or <1200 *Ascospores/Basidiospores* /m^3 constitutes clean by their definition.

EPA is researching ways to remediate mold-contaminated buildings without gutting them by testing the ability of chlorine dioxide gas to render fungi, their fragments, and their mycotoxins harmless. Such work generally requires DNA analysis using PCR. Work at Texas Tech (Wilson et al 2005) has shown that some fungi can be inactivated, but others remain toxic.

Testing and Reporting the Efficiency of In-Duct UV Light Technologies

Testing a UV light device installed in a test rig that represents the ductwork of an HVAC system is generally straightforward. The device is challenged with bioaerosols and the air is sampled upstream and downstream of the device getting 6 to 12 samples for each measurement. Each sample is plated and grown out and the CFUs are counted. The counts for all the measurements are summed and the penetration and efficiency calculations are performed using the sum. A problem occurs when the sum of the counts is zero, as occurs when the microorganisms are very efficiently killed. Non-detects do not necessarily mean zero microorganisms, therefore 100% efficiency for the device is not reported when the outcome of downstream tests is *complete non-detection* of organisms. How then should the calculation be handled?

Restructuring the test to get a few counts downstream should be considered, however providing more bioaerosol concentration upstream generally will not affect downstream counts and neither will longer downstream sampling times or plating more sample. The solution is to consider that these low downstream numbers have a Poisson distribution and to use a Poisson statistics table to provide a number that is an upper bound for any sum of counts less than 50 (including zero). Poisson statistics have the unique property that the sum of Poisson averages is still Poisson. Thus an efficiency can be reported that is "greater than" the calculated value, generally >99%.

Conclusions

The traditional thought processes that govern planning for sampling, analysis, and detection limits of microorganisms are somewhat different from those for non-biological substances. To analyze risk the sampling plan should consider probability as well as targeted sampler siting. Both viable and non-viable microbes and their products should be collected and analyzed. Research on molecular methods is improving the science of identification of microorganisms. The human element is often as important as the science. For people to be comfortable living or working in a building after a contamination and subsequent decontamination event, stakeholders must be drawn in and educated at the outset and throughout the process.

References

ANSI/ASHRAE (American National Standards Institute/American Society of Heating, Refrigerating and Air-Conditioning Engineers). 1999. ANSI/ASHRAE Standard 52.2-1999, Method of testing general ventilation air-cleaning devices for removal efficiency by particle size, Atlanta, GA

Baxter DM, Perkins JL, McGhee CR, and Seltzer JM. (2005) A regional comparison of mold spore concentrations outdoors and inside "clean" and "mold contaminated" Southern California buildings, J Occup Environ Hyg. 2:8-18

CDC (Centers for Disease Control). 2002 Comprehensive procedures for collecting environmental samples for culturing bacillus anthracis
www.bt.cdc.gov/agent/anthrax/environmental-sampling-apr2002.pdf

CDC 2005. Mold prevention strategies and possible health effects in the aftermath of hurricanes Katrina and Rita
www.bt.cdc.gov/disasters/mold/report/pdf/2005_moldreport.pdf

CRS (Congressional Research Service.) 2005. Cleanup after hurricane Katrina: Environmental Considerations, CRS Report for Congress, October 13, 2005, Congressional Research Service, the Library of Congress, Order Code RL33115

Dun S, J Wood, and Martin B. 2005. Workshop on decontamination, cleanup, and associated issues for sites contaminated with chemical, biological, or radiological materials, EPA Report, EPA/600/R-05/083

40 CFR part 136, Appendix B (1984)

Definition and Procedure for Determination of the Method Detection Limit

GAO (Government Accountability Office). 2005. Anthrax Detection: Agencies Need to Validate Sampling Activities in Order to Increase Confidence in Negative Results, GAO Report, GAO-05-251, www.gao.gov/cgi-bin/getrpt?GAO-05-251

IICRC (Institute of Inspection, Cleaning and Restoration Certification). 2003. IICRC S520 Standard and Reference Guide for Professional Mold Remediation, Vancouver, WA

Macher, J. (ed.) (2005) Bioaerosols: Assessment and Control, American Conference of Governmental Industrial Hygienists (ACGIH) Cincinnati, OH

NAS (National Academy of Sciences). 2005. Reopening Public Facilities after a Biological Attack: A Decision-Making Framework, Committee on Standards and Policies for Decontaminating Public Facilities Affected by Exposure to Harmful Biological Agents: How Clean is Safe? National Research Council, ISBN: 0.309-09661-8, National Academy of Sciences, www.nap.edu/openbook/0309096618/html/168.html

USEPA (U.S. Environmental Protection Agency). 2001. Mold remediation in schools and commercial buildings, EPA Report, EPA 402-K-01-001

Wilson SC, C Wu, LA Andriychuk, JM Martin, TL Brasel, CA Jumper, and DC Straus, (2005) Effect of chlorine dioxide gas on fungi and mycotoxins associated with sick building syndrome. Applied and Environmental Microbiology, 71:5399-5403

When Laboratories Should Not Censor Analytical Data, and Why[1]

Charles B. Davis[2] and Nancy E. Grams[3]

Abstract

Analytical laboratories conventionally censor measurements below a "Reporting Limit", reporting them as "nondetect" or "<RL". While doing so is based on sound scientific principles for interpreting individual measurements, it is counter-productive when decisions are being made from entire datasets and/or when the data user desires a reliable detection estimate for a project.

Both of these needs arise in facility surveys for worker protection from surface contamination. A common industrial hygiene approach to "protecting all workers" is to compute an Upper Tolerance Limit (UTL = upper 95% confidence limit for the 95th percentile) from data obtained according to a scientifically sound sampling plan, and declare the facility "safe" if the UTL is less than a Regulatory Criterion (RC).

UTLs take two forms: parametric, available when one may assume a known distribution family (almost always normal or lognormal) and the censoring proportion is not too large; and nonparametric, available with any distribution or censoring proportion. The latter requires at least 59 samples to achieve the desired confidence, whereas the parametric approach can make substantially more efficient use of data.

For example, as part of a large-scale facility survey 14 swipe samples were taken in an office trailer on the Nevada Test Site. Uncensored beryllium measurements were obtained. The parametric UTL based on these data is well below the RC; the facility is clearly "safe". But if these data had been laboratory-censored using an RL one-tenth the RC, all data would be nondetects, and one could not decide "safe" with fewer than 59 samples. Ironically, with the same RL, if all 14 measurements were three times higher, the lower nondetect proportion would allow using a parametric UTL, which would be below the RC. The appropriate "safe" decision could be reached with only 14 samples if the trailer were three times more "contaminated" than it is, whereas if the original data had been censored over four times as many samples would be needed.

A related issue concerns the Method Detection Limit (MDL) as implemented in 40 CFR Part 136 Appendix B. In the facility survey setting, in addition to the usual facets of the on-going controversy over the MDL, one may also question the propriety of using MDLs obtained from non-blind analyses of spiked reagent water to provide insight into the interpretation of analyses of project samples (digested swipes). Rather than relying on this questionable source of information, if one absolutely must have an RL and double-blind project blank

[1] The views expressed in this article are those of the authors, and are not intended to represent policy of the Nevada Test Site, Bechtel Nevada Corporation, or any Department or Agency of the U.S. Government.

[2] Principal Statistician, EnviroStat, Las Vegas, NV; charles.davis@envirostat-nv.com.
[3] President, Advanced Earth Technologies, Elgin, IL.

samples are available, one should return to Lloyd Currie's original, simpler L_C concept, using uncensored blank measurements. Doing so would require that no data be censored by the laboratory.

Nearly five centuries ago a keen observer of human nature cautioned his pupils about the perils of attempting to modify an accepted *status quo*.[4] The conventional laboratory practice of censoring observations below a Reporting Limit, reporting them as "nondetects" or as "<RL," is one such accepted practice that promises to be resistant to change. Nonetheless, for providing data for decision-making in facility surveys, this practice is both poorly grounded in scientific principle and counter-productive.

Instead, **laboratories should avoid censoring data to be used in facility surveys** whenever and wherever possible[5]. Rather, they should provide uncensored data, along with their conventional RLs. This recommendation applies whenever decisions are to be made through the statistical evaluation of an entire data set rather than from individual measurements. Uncensored data, including double-blind performance evaluation data, could also be quite valuable to data users in other circumstances, such as determining reliable project critical levels (see L_C to follow) and project-level detection levels (see L_D to follow) and in assessing the attainment of data quality objectives.

Statistical Principles Underlying Detection Limits

The broadly accepted conceptual approach to detection was developed by Lloyd Currie (1968). Lloyd's Limits[6] are the following:

L_C = Critical Level = the minimum measurement statistically distinguishable from the distribution of signals with no analyte present, to be used in determining whether the analyte is detected; and

L_D = *a priori* Detection Limit = the minimum true concentration above which a measurement
method will consistently detect the analyte at a defined level of statistical confidence
(using L_C for deciding whether a detect is found).

Currie's concepts as presented are clean and attractive. Implementing them in scientifically sound ways, taking into account practical limitations on available data

[4] "There is nothing more difficult to take in hand, more perilous to conduct, or more uncertain in its success, than to take the lead in the introduction of a new order of things." *The Prince*, Niccolò Machiavelli, 1532.
[5] Some analytical methods, such as GC/MS for organics, identify nondetects using both quantitative and qualitative criteria, whereas with other methods, such as ICP-AES for metals, a measurement can always be determined. Even for the former, project information can be greatly enhanced by applying these concepts to the quantitative aspect of the determination of nondetects.
[6] Currie also defined L_Q = Determination Limit = the minimum true concentration above which individual measurements have a desired precision. L_Q is not involved in facility survey decision-making.

quantity and quality and resulting uncertainties in calibration functions and other estimates of analytical response, has proven elusive. In particular, although the Method Detection Limit (MDL) as codified in 40 CFR Part 136 Appendix B was originally promoted as implementing the L_D concept (see Glaser *et al.* 1981), U.S. Environmental Protection Agency (EPA) has recently acknowledged that it should more nearly be associated with L_C (see U.S. EPA 2003). Part of the difficulty may stem from the urge to use L_D and/or the MDL as a reporting limit, since the word "detection" appears in its name, even though L_D clearly can not have its defining property unless L_C is used as the detection threshold! See Grams (1990) for a detailed discussion of MDLs and ASTM (2003) for a rigorous implementation of Currie's L_D concept.

Most importantly, Lloyd's Limits were developed for interpreting individual analytical results. In facility surveys one rarely makes a decision using an individual measurement, unless that measurement itself exceeds a Regulatory Criterion (RC). Rather, in most situations the entire data set obtained following a scientifically designed sampling plan is compared statistically with the RC. Currie's concepts were not designed for and should not be applied to measurements used in this way.

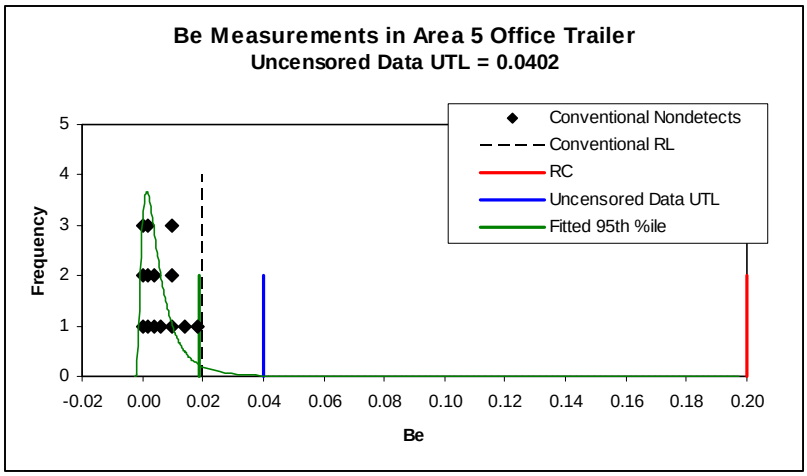
The Adverse Impact of Censoring Data to be Used in Facility Surveys

This discussion would be largely academic were it not for the adverse and perverse effect that censoring data has on statistical decision-making in facility surveys. For example, surface beryllium contamination is currently a significant concern within the U.S. National Nuclear Security Administration, Department of Energy (NNSA). The applicable regulation is 10 CFR 850; its intent is generally interpreted to require that "no worker" should be "exposed" to removable surface Be concentrations exceeding the RC of 0.2 µg/100 cm^2. Current Industrial Hygiene (IH) practice implements this by requiring that an upper 95 percent confidence limit for the 95^{th} percentile of the distribution of measurements in 100 cm^2 swipe samples obtained at a facility should not exceed the RC. In statistical jargon, this limit is an Upper Tolerance Limit (UTL) with 95 percent confidence and 95 percent content. (See the Appendix discussion.)

There are two common ways to compute UTLs. The nonparametric UTL is simply an upper order statistic of the data distribution. One needs at least 59 measurements to compute a nonparametric 95 percent-95 percent UTL. This UTL is the highest observation if the sample size N is between 59 and 92, the second highest observation if N is between 93 and 123, the third highest if N is between 124 and 152, and so on. Following the nonparametric approach one obtains 59 measurements, and if all are below the RC the facility is declared "safe."

The other common approach is based on assuming a known distribution family, almost always normal (Gaussian). On an appropriately transformed scale (log, usually) one computes UTL = (sample mean) + K * (sample standard deviation), K being a tabled value (see, e.g., Gilbert 1987). UTL is then re-expressed on the original scale for comparison with the RC. Parametric UTLs can be computed when some of the data are censored. Doing so reliably is an area of on-going statistical research; see Davis (2006),

for example. The nondetect proportion should not exceed 70 percent using currently available censored-data UTL methods.

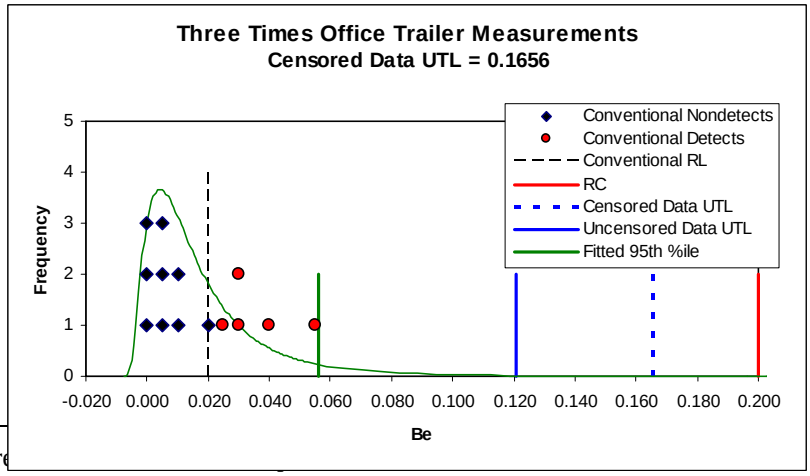


Parametric UTL techniques can use data considerably more efficiently than the nonparametric procedure. The sample size needed depends on how low one expects the data distribution to be relative to the RC. If the mean is expected to be several standard deviations below the RC (on the transformed scale), as would be anticipated in many situations, considerably smaller sample sizes than 59 will often be adequate.

As an example, consider these N = 14 observations from an office trailer located in Area 5 of the Nevada Test Site[7] (NTS). The laboratory was asked to provide uncensored values; it did so, along with a disclaimer stating that doing so is not typical practice and a statement of what the typical RL would be. Using uncen-sored data the UTL is 0.0402 µg/swipe, which is well below the RC. This facility is clearly "safe." (See the Appendix for details and a data listing.)

When data are laboratory-censored, data sets with more very low values end up with higher proportions of nondetects. For purposes of illustration, suppose that the laboratory had censored the data in this study using an RL of 0.02, which is only one-tenth of the RC. All of the data would then be nondetects. Accordingly, the conventional wisdom would say that the normal-distribution UTL could not be used, so one would have to use the nonparametric approach, which would require augmenting the sample size over four-fold to 59 before making a decision for this facility.

But consider the same situation (same trailer, same laboratory, same RL) with higher data values. If each observation were three times higher, the laboratory would report five numerical results and nine nondetects (64 percent nondetects),



[7] The authors express their appre

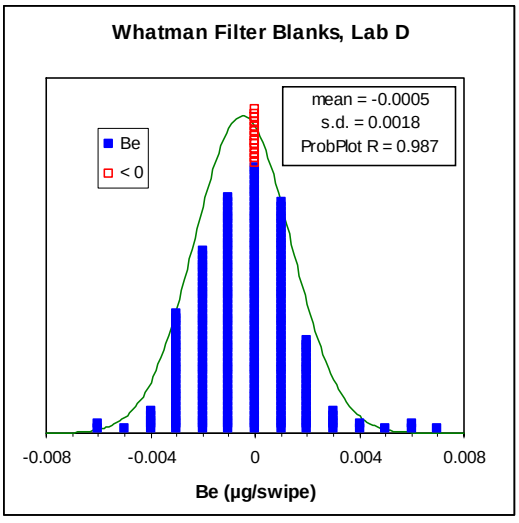
the normal-theory parameteric UTL could be applied (UTL = 0.1656, below the RC), and the facility would be declared "safe." The irony, of course, is that the conven-tional procedure would allow one to conclude that the office trailer is "safe" with only N = 14 measurements if they were three times higher than they actually are, but not with their actual values if those values were censored following current common practice!

Clearly, changing conventional practice to avoiding laboratory censoring of data in these types of projects is important. When data are laboratory-censored, arriving correctly at a "safe" decision becomes less efficient as the measurement levels decrease, contrary to what one should expect. There is perhaps a parallel between this recommendation and that arising in the U.S. EPA's Triad Approach (Crumbling *et al.* 2001). The common theme is to make data quality considerations at the individual measurement level subservient to the efficiency and accuracy of the overall decision-making process.

A Better RL, In Case One Absolutely Needs an RL

The common laboratory censoring limit in use in the U.S. is the MDL, based on Glaser *et al.* (1981). Although those authors state an intention resembling that of Currie's L_D, they make use of questionable "simplifying assumptions" that do not appear to be supported in our observations of uncensored Be data. Moreover, it is difficult to reconcile either their technical discussion or the use of the MDL as a censoring value with either L_D or L_C; see Grams (1990). U.S. EPA (2003) recently clarified that the MDL should be associated conceptually with Currie's L_C. Accordingly, one should no longer associate the MDL with concepts such as "the lowest value of an analyte that a measurement method will consistently see with a defined level of statistical confidence," a concept properly associated with L_D, not L_C.

There are additional conceptual difficulties with using MDLs to implement Currie's L_C concept. One is that MDLs are determined from off-line analyses, often conducted with an eye on contractual requirements that must be met, and MDL determinations can be influenced by the choice of spiking concentrations. The spiking concentrations do not include zero. These factors can weaken the hoped-for relationship between MDLs and the actual statistical properties of routine environmental measurements. Another is that MDLs are estimated from spiked samples prepared in liquid, typically water, and it is not clear how these are related to the statistical properties of measurements made on digested swipe samples.



But suppose that one simply must have an RL that functions reliably as Currie's L_C. There is a good candidate, so long as one has uncensored data from double-blind-to-the-laboratory blank samples; one can simply

return to Currie's original L_C itself. The most direct and arguably most correct approach is to take the RL to be an upper 99 percent prediction limit (UPL) obtained from the distribution of blank measurements. Any measurement above that is by definition statistically distinguishable from the distribution of blank measurements, so the UPL becomes the project L_C. The laboratory must not know which samples are the blanks, to preserve double-blind quality control. This implies that no measurements should be censored by the laboratory. Rather, laboratories should provide uncensored data, and then state separately what their conventionally determined reporting limit would be, along with the details of its derivation.

As an example, the preceding plot shows the distribution[8] of 286 field blank observations obtained using the same swipe medium and analyzed by the same lab during a seven-month period, including the time the NTS office trailer swipes were obtained. L_C determined as suggested above would be 0.0038; this agrees with the lab's stated RLs of 0.003 to 0.005. It is worth noting that when the sampling media were switched from Whatman filters to Ghost Wipes, the lab's stated RLs did not change, but L_C computed from blank data increased to around 0.014!

Laboratory Performance Evaluation and Data Quality Objectives

U.S. EPA (2003) suggests that the MDL procedure provides a means of assuring that laboratories can provide analyses of adequate sensitivity for their intended purposes. For this reason, laboratories should continue to report their conventionally determined MDLs. Again, however, MDLs are not based on double-blind measurements and often do not involve the actual sample medium to be used. Accordingly, although MDL values should continue to be requested and reported, they should not be the only means of performance evaluation. In addition, facilities should submit double-blind performance evaluation samples to their labs. Spiking levels and evaluation criteria should be developed consistent with the aim of ensuring data of appropriate quality for their efficient (uncensored) use.

A more challenging issue is that of providing legitimate ways to establish and implement Data Quality Objectives (DQOs) for future studies. With the MDL having been realigned conceptually with L_C, one must find an alternate implementation of the L_D concept to meet this challenge. Space does not permit an extended discussion of this topic here, but two essential issues related to the theme of this paper are the following: (1) developing an implementation of the L_D concept over an appropriate inter-laboratory scope; and (2) developing protocols for assessing whether or not a laboratory not yet under contract might be able to meet the DQOs so developed. These issues are discussed in ASTM (2003).

[8] Due to an oversight the database initially stored negative values as zeros; this was subsequently corrected. Censored data MLEs of mean and standard deviation are given; see the Appendix.

Summary and Recommendations

In summary, current laboratory practice regarding development and use of reporting limits is well established and is likely to be resistant to change. Nonetheless, this practice, developed for evaluating individual measurements, is unsound in principle and counter-productive in result when applied to samples obtained for use in facility surveys and similar situations. The following recommendations are distilled from the preceding discussion:

- Laboratories should avoid censoring data to be used in multiple-measurement decisions to the extent possible. If laboratories must censor data, the data user should require the lowest RL available. When reporting uncensored data, the laboratory should state its MDL or other conventionally determined RL separately, along with the details (method, spiking levels, and data) of the RL determination.
- Doing so will allow facilities to benefit from the efficiencies available from using parametric statistical analyses for demonstrating regulatory compliance. A major benefit is that as measured values decrease, one can demonstrate compliance with fewer observations, which is not the case when data are censored by the laboratory as is currently common practice.
- Facilities wishing to use a reporting limit with swipe sample measurements should determine their own RL as a 99 percent upper prediction limit computed from the distribution of measurements of uncensored double-blind field blank samples. Doing so avoids the conceptual difficulties inherent in the MDL implementation in 40 CFR Part 136 Appendix B and furthermore ensures that the RL will be appropriate for swipe samples and matrices.

References

ASTM (2003). *Standard Practice for 99%/95% Interlaboratory Detection Estimate (IDE) for Analytical Methods with Negligible Calibration Error*, D6091-03, ASTM International.

Currie, L.A. (1968). "Limits for Qualitative Detection and Quantitative Determination: Application to Radiochemistry", *Analytical Chemistry* 40, 586-593.

Crumbling, D.M., C. Groenjes, B. Lesnik, K. Lynch, J. Shockley, J., Van Ee, R. Howe, L. Keith, and J. McKenna (2001). "Managing Uncertainty in Environmental Decisions", *Environmental Science & Technology*, October 1, 404A-409A.

Davis, C.B. (2006). "Parametric 95%-95% Upper Tolerance Limits for Censored Lognormal Data", preprint available on request.

Gilbert, R.O. (1987). *Statistical Methods for Environmental Pollution Monitoring*, Van Nostrand Reinhold, New York.

Glaser, J.A., D.L. Foerst, G.D. McKee, S.A. Quave, and W.L. Budde (1981). "Trace Analyses for Wastewaters", *Environmental Science & Technology* 15, 1426-1435.

Grams, N.E. (1990). "A Statistical Approach to Calculation of the Method Detection Limit and Its Application to Volatile Organics by GC/MS", presented at the 1990 Water Quality Technology Conference, November 11-15, San Diego; preprint available on request.

U.S. EPA (2003). "Guidelines Establishing Test Procedures for the Analysis of Pollutants: Procedures for Detection and Quantitation", 68 *Federal Register* 11770-11790, March 12. See also "Technical Support Document for the Assessment of Detection and Quantitation Concepts", 68 *Federal Register* 11791-11793, March 12, and Appendices referenced therein.

Appendix

The Appendix contains technical details and data listings. It is available via e-mailed request.

A New Statistical Procedure for Determining Minimum Reporting Levels Under the Unregulated Contaminant Monitoring Regulation (UCMR 2)

John J. Martin*, Stephen D. Winslow, and David J. Munch

The unregulated contaminant monitoring program was designed by EPA to collect monitoring data for contaminants suspected to be present in drinking water, but that do not have health-based standards set under the Safe Drinking Water Act (SDWA). Every five years the list of contaminants is revised largely based on the Contaminant Candidate List (CCL). The current program is managed as a federal EPA direct implementation effort. For the second cycle of the program (UCMR 2), all laboratories providing UCMR 2 analysis are required to demonstrate their ability to measure each specified compound at or below the Minimum Reporting Level (MRL) for that compound.

The UCMR program was designed to produce uniform national drinking water contaminant data for a targeted list of contaminants of concern. Uniformity of data will help to reduce instances of under-reported contamination, which may have consequences concerning public health; and help to minimize false reports of contamination, which may trigger an assessment that places an unnecessary economic burden on drinking water utilities. Acceptable levels of both precision (i.e., the reproducibility of the data) and accuracy (i.e., how close measured values are to the true value) are very important to the overall implementation of this regulatory program. To this end, EPA has developed a process for determining the single-laboratory Lowest Concentration Minimum Reporting Level (LCMRL). The LCMRL is the lowest true concentration for which the future recovery is predicted to fall, with high confidence (at least 99%), between 50 and 150% recovery. This recovery interval has been used for the past several years in occurrence data gathering efforts, including the Information Collection Rule (ICR) and the first cycle of UCMR (UCMR 1).

The process of analyzing an environmental sample for the presence of contaminants includes the determination of a detection limit and a quantitation level. Detection and quantitation procedures have been the subject of a great deal of research for several years. EPA has established a Federal Advisory Committee on Detection and Quantitation Approaches (FACDQ) which is considering the formal adoption of procedures by EPA. Although the LCMRL was developed for use in UCMR 2, the procedure is flexible in terms of the selected confidence level and recovery criteria, thus allowing it to be applied to any analytical regulatory program.

LCMRLs are laboratory-specific and were calculated by selected laboratories as part of method development for the UCMR program. LCMRLs were used to establish MRLs for analytes specified in the recently-promulgated UCMR 2; hence, MRLs exhibit the same Data Quality Objectives (DQOs) as those presented for LCMRLs. The LCMRL/MRL are quantitation levels that do not address the issue of detection; however, they do address the potential for non-constant variance over the range of replicate spiking concentrations. The process for determining LCMRLs and MRLs is presented along with a discussion of the implications of the MRLs on participating laboratories.